Text Analysis with a #ZineCorpus

Filipa Calado @Caladoscope, Jenna Freedman @zinelib, and Miranda Johnson present at #NYCDHWeek22

Please feel encouraged to enter your name, location, and reason(s) for attending this session in the chat! & download the zip file.

The Corpus

#ZineCorpus

Introduction

Filipa (she), Miranda (they), Jenna (she)

Thank yous

- Columbia Libraries Ask A Librarian
- Alicia Peaker
- All of the zine creators who said yes!

What this session is

- Insight into a work in progress
- Invitation to experiment respectfully with zines in text analysis
- introductory

What this session isn't

- Using python despite what we said in the description sorry
- Ready for distant reading
- Complete

Questions/discussion: type into chat

#ZineCorpus

Building the Corpus: Motivations and Goals

- Experimenting with how zines might fit into DH work
- Desire to see zines studied as sources of information, not just objects
    - Big Question: what is gained and lost with step away from physical/visual object?
- Outreach— get more people reading zines!

#ZineCorpus

Building the Corpus: September 2021

- A bit about me
- Basic project concept already outlined
  - Corpus of plain text transcriptions
  - Some access point/place for metadata
- Already decided to focus on previously digitized zines
- Experimentation with Voyant

#ZineCorpus

Building the Corpus: Pilot Phase

- 5 zines chosen for range of format, genre, author identity
- Creating the catalog
  - Modified xZINECOREx
  - Added author identity access points
  - Linked sheets system
- Communication around consent and permissions
- Began transcription

#ZineCorpus

Building the Corpus: Pilot Phase— What We Learned

- Author identity set up was WAY too complicated
  - Eliminated linked sheet system
- Moved to opt-out language for second phase
- Transcription learning curve
  - Transcribing *for text analysis*
  - OCR and time estimates

#ZineCorpus

Building the Corpus: Transcription Phase

- Selecting Zines
  - Attempting a representative sample
- Transcription process
  - OCR
  - Copy and Paste by section
  - Clean up
- Continuing to play with representing identity categories

#ZineCorpus

Building the Corpus: Decision Making

- New corpus format
- Technical vs. Ethical considerations
  - No easy split for this project!

#ZineCorpus

Building the Corpus: Decision Making

Permissions

- Striving for clear but casual, friendly, and concise

Miranda Johnson <mj3064@columbia.edu> Dec 17, 2021, 6:12 PM
to contact, zines

Hi SWARM Collective members,

I'm Miranda Johnson, and I'm writing to you on behalf of the Barnard Zine Library as an intern working on our Zine Corpus project. With your permission, we would like to include the 2014 zine *Femme whore*, written by SWARM members for Sex Worker Open University, in the project.

We are in the beginning stages of creating the zine corpus, which will be a dataset of sharable zine text files that could eventually be used for digital humanities textual analysis. The corpus itself will consist of a spreadsheet with links to the plain text files (including your zine!). The goal of this little PDF → text file conversion project is to facilitate research by scholars, researchers, zinesters (inside and outside of Barnard) and the public, embedding zines more deeply into digital scholarship. We want to encourage people to consider zines as sources of information in their own right, rather than just as objects, and zinemakers as active participants in important conversations.

Femme whore is in our collection and it's also available as a PDF on our Barnard Zines with Online Links list, which is why we'd like to include it. **Please let us know if you'd like to opt-out of our zine experiment!**

Researchers in our library are often interested in race, ethnicity, gender, sexuality, etc. in zines, so we are including descriptors in the spreadsheet to make those identities more visible. We will only include information that is in the zine itself or that you want to provide to us.

Several of us are zine makers or hold one or more minoritized identities ourselves and we intend to proceed with close attention to the needs of the creators whose works comprise our corpus. Have a great day, and feel free to ask us any questions!

Warmth and Solidarity,
Miranda
(they/them)

Building the Corpus: Decision Making

Permissions

- Opt-in vs. **opt-out**?
  - Previous permissions
  - Quick timeline
- No opt-outs, several non-responses
- Cases of shaky contact info
  - collectives/anthologies?

Building the Corpus: Decision Making

Representing Creators' Identities

- Anticipated author identity markers as major access points
- Creator-preferred language vs. standard vocabulary for ease of access
  - Include both creator term and "umbrella group", linked sheets
  - Alicia Peaker's reminder: "corpus, not a database!"
  - Umbrella terms → marginalized/not marginalized/ none given

#ZineCorpus

Building the Corpus: Decision Making

Representing Creators' Identities

- ID markers for anthology zines
  - **Inclusive** or exclusive
  - But not very clear in catalog
- How much inference to do? Outside research?
  - Decided on none

#ZineCorpus

Building the Corpus: Challenges

- Approximating a representative selection. Biases toward:
  - Recent, born digital
  - OCR compatible - PDF and typed vs. handwritten
  - Interested me
- Author contact info and response times
- My lack of digital humanities background
  - harder to predict user needs

#ZineCorpus

Building the Corpus: Going Forward

Re-evaluate?

- Current creator identity model collapses racial categories
  - Especially hard to square when some zines are explicitly pushing back against this (e.g. *Our Sea of Islands)*
- Similarly, gender category currently includes both women and trans people of all genders

#ZineCorpus

Building the Corpus: Going Forward

But mostly…

NEEDS MORE ZINES!

#ZineCorpus

Exploring Tool Paths

Choose one word from the 10 most frequent words in the corpus (check the Cirrus or Terms pane).

Then, choose two or three texts in the corpus where this word is used very frequently (check the titles on the X-axis in the Trends pane).

Finally, explore how this word appears within these different texts in the corpus. You might consider:

- What other words appear in close proximity with this word? (TermsBerry, Links, Collocates)
- How does the usage of this word compare with other frequent words within each text (Trends, Bubblines)?
- While using the tool, consider the benefit for analysis. How might this tool help to identify an area of research?
- Or, just play around!

Remember: To change the tool, click the four dark rectangles at the top right of each pane. On most panes, you can “clear” or “reset” the data if you want to explore a different word. And on some of them, you can directly type or input the word that you want to explore.

#ZineCorpus

Exploring Individual Tools

In the blue navigation bar at the top of the screen, click on the four little rectangles on the top right.

Under “Corpus Tools”, choose one of the tools listed, such as “Loom”, “MicroSearch” “Dreamscape” or “Veliza”.

When the tool is open, explore some of the options for configuring the tool on the bottom right.

While using the tool, consider the benefit for analysis. How might this tool help to identify an area of research?

Or, just play around!

Remember: To change the tool, click the four dark rectangles at the top right of each pane. On most panes, you can “clear” or “reset” the data if you want to explore a different word. And on some of them, you can directly type or input the word that you want to explore.

#ZineCorpus

Discussion

We'll take a **progressive stack** and ask you to keep these questions in mind (per Eve Tuck on Twitter):

a) make sure it is really a question;

b) make sure you aren't actually trying to say that you should have given the workshop;

c) figure out if the question needs to be posed and answered in front of everyone;

#ZineCorpus

Follow Up

Filipa fcalado@gc.cuny.edu

Miranda miri.o.johnson@gmail.com

Jenna jfreedma@barnard.edu

#ZineCorpus